DOCUMENT RESUME

ED 220 499 TM 820 512

AUTHOR Cole, Nancy S.
TITLE Grade Equivalent Scores: To GE or Not to GE.
PUB DATE Mar 82
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Elementary Secondary Education; Evaluation Methods; Formative Evaluation; *Grade Equivalent Scores; *Measures (Individuals); Scores; *Scoring Formulas; Testing Problems; *Test Interpretation
IDENTIFIERS *Grade Development Scale

ABSTRACT

The advantages and disadvantages of grade equivalent (GE) scores are explored, including appropriate uses for GE type scores and how to bring current GE scales closer to the type of information educators appear to desire. Although GE scores are not an equal interval scale, not comparable across school subjects, and do not indicate the grade level at which a student can perform, the GE scores are suitable to identify student grade level skills and to chart growth over time in the "what skills" sense, not the "how much growth" sense. Given these potentially valuable uses, a variant of a GE scale that could more nearly describe the grade level at which students are performing is introduced as the Grade Development Scale (GDS). A GDS score produced less discrepant estimates of grade level from current grade for extreme scores than any of the common GE scales. The scale is not an equal interval scale, nor should it be used in comparing subject areas. Problems of interpreting the amount of growth in GDS units remain. Studies are underway to determine if the GDS scores will be found useful to educators in estimating the grade level at which a student is performing. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED220499

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☒ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

GRADE EQUIVALENT SCORES:

TO GE OR NOT TO GE

by

Nancy S. Cole

University of Pittsburgh

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
N. S. Cole
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

March 1982

TM 820512

Grade Equivalent Scores: To GE or Not to GE[1]

Nancy S. Cole
University of Pittsburgh

Grade equivalent scores (GE scores) have long been a part of standardized school achievement testing. They have undoubtedly been popular because they appear to give the desirable information of which grade level skills a student can perform. At the same time, GE scores have been the target of many criticisms. Their popularity suggests there are favorable aspects to GE scores--some potential advantages to their use--and the criticisms are statements of their unfavorable aspects, their disadvantages.

The purpose of this paper is to begin the task of sorting out and exploring these advantages and disadvantages. The first section identifies many of the criticisms that have been made of GE scores although it is not intended as an exhaustive analysis of those criticisms. In the second section we note some major uses of achievement test results and the types of score scales designed for different uses. From this perspective, many of the criticisms of GE scores can be seen as recommendations that other types of score scales would be more appropriate than GE scores for certain uses. In this section we also note the fundamental use that seems to be the appropriate province of GE-type scales and the remaining criticisms that relate to this use. In the third section we propose a method to produce a GE-type scale, called a Grade Development Scale (GDS), that appears to better accomplish the fundamental

[1]AERA Division D Vice Presidential Address, American Educational Research Association Annual Meeting, New York City, March, 1982

TM 820 512

purpose of GE-type scales—to describe the grade level at which a student performs.

The reader should be forewarned that the result of this analysis is not an answer to the question as posed, "To GE or Not to GE?" Rather than conclusions about whether or not to GE, this exploration produces thoughts about when and how to GE and not GE. Perhaps it is fitting that when starting from Hamlet's famous question, we, like Hamlet, see complexities in the real world that force us away from the simple question to more analytical and typically academic qualifiers of the situation.

Criticisms of Grade Equivalent Scores

As noted, there have been many criticisms made of GE scores by many authors. Here we group the criticisms into three categories that seem to cover the major areas the specific criticisms address.

Not An Equal Interval Scale

It is commonly noted that GE scales are not equal interval scales. This fact is stated in many test manuals (e.g., Prescott, Balow, Hogan, and Carr, 1978), noted in measurement texts (e.g., Lindvall and Nitko, 1975; Thorndike and Hagen, 1977), and frequently cited as a criticism of GE scales (e.g., Horst, 1976; Berk, 1981; Linn and Slinde, 1977). The basic concern here is that one GE unit does not have the same meaning in different parts of the GE scale—one year's or one month's growth may represent very different amounts of growth at different points in the scale. Closely related are concerns (1) that expected normal growth is not the same for students at different score levels—a low scoring student may have a normal expected growth in one year of .7 GE units whereas for an average scoring student it is 1.0 GE units and for a high scoring

student the expected growth may exceed 1.0 GE units (Linn, 1981)--and (2) that GE curves smooth uneven growth rates within the school year and do not accurately reflect the complexities of summer gains or losses (Bark, 1981).

The assumption of at least approximate equal interval properties is needed to add the scores of individuals together to obtain a group mean and to interpret the amount of gain for individuals meaningfully. These properties are especially important for evaluation studies of instructional effects or research studies involving academic growth over time.

Not Comparable Across School Subjects

There is common agreement among test makers and other writers that a grade equivalent score from a test in one school subject (e.g., reading) should not be compared to a GE score from another subject (e.g., mathematics) to try to infer an individual's (or group's) relative strength in the two areas (Hieronymus, Lindquist, and Hoover, 1978; Lindvall and Nitko, 1975; Prescott et. al., 1978; Thorndike and Hagen, 1977). The problem here is that because growth rates in different subjects are different, two scores with identical relative standing within a grade group (e.g., scores in reading and math both with a percentile rank score of 95) may produce different GE's. It is generally agreed that it is more reasonable to conclude that a student with the same percentile rank in two subjects is performing with similar strength in the two areas than to conclude different strengths based on GE scores.

Does Not Indicate Grade Level at Which Student Can Perform

Thorndike and Hagen (1977) note that for an advanced third grader who obtains a GE of 5.9, "This does not mean that this child has a mastery

of the arithmetic taught in the fifth grade... [or] ... that this child is ready to move ahead into sixth grade work" (p. 120). What such a score probably means is that this child has learned third grade arithmetic very well (Lindvall and Nitko, 1975) but it does not tell directly what levels above third grade the child has actually learned. This deviation from the apparent meaning of the score occurs because GE scales are constructed to give the grade level at which other students score the same as this student on the content tested, and late-year fifth graders apparently scored the same as this student on the third grade arithmetic test taken. GE scales do not directly inform us about what level of content above the third grade this high scoring student can actually perform as again critics and proponents of GE's agree. It is this feature that led Linn (1981) to note that "Much of the popularity as well as much of the criticism stems from a common characteristic, namely, the fact that the GES invites seemingly simple but misleading interpretations" (p. 92).

The inappropriateness of the desired interpretation (the grade level at which a student is performing) is made even clearer by examining the GE scores achieved on some tests for high- and low-scoring examinees. It likely does not seem reasonable to anyone to give a third grader a GE of 6.0 or 7.0 or even 8.0 when teachers, parents, and the students themselves realize such students have never even been exposed to the content of these far advanced grades (Angoff, 1971). Extreme scores such as this are common on some GE scales and make it clear to most observers that such scores could not be describing the grade level work a student can perform.

Related to this interpretive concern is the concern that even the notion of "performing at a grade level" has questionable meaning beyond the elementary years (Thorndike and Hagen, 1977), and as a result, GE scales have even less desirable properties at the high school level (Berk, 1981; Reynolds, 1981). As further evidence of the discrepancy of GE scores from the desired simple interpretation users might wish to make, many have noted the differences across publisher and major batteries in the characteristics of the GE scale (Berk, 1981; Lindvall and Nitko, 1975). A student scoring at the same within grade percentile rank on two different tests may receive widely differing GE scores. Thus, even if the GE scale from one test produced scores that could be interpreted in terms of the grade level at which a student is performing, clearly all the other tests could not be producing such scores.

Such discrepancies across tests and publishers bring into question as well the reasonableness of the widely-used "two years below grade level" as a criterion for special treatment regardless of test and publisher (Reynolds, 1981). When the GE scale difference from test to test is combined with the lack of equal interval property so that two years below grade level means different amounts below grade level for different grades and subjects, it is clear this use deserves attention.

A Look at Score Types and Uses

Many users have undoubtedly wondered why tests have to have so many different types of scores. What a maze it must be for the uninitiated to confront the choice between percentile ranks (PR's), stanines, normal curve equivalents (NCE's), grade equivalents (GE's), and variously-named expanded standard scores! Even measurement experts

seem, at times, to have forgotten the reason for this maze. That reason is, of course, that different scores with different properties are suited to some uses and not to others so several types of scores are given so the appropriate one can be chosen for the particular use. We do not average percentile ranks because they are not suited to that but we do use them as the fundamental information about an individual's standing in a group. We need to examine the criticisms of grade equivalent scores noted above from the same perspective.

Horst (1976) argued that "in general, the whole point of obtaining a test score is to ascertain either

(a) what skills a student has, or

(b) where the student stands in relation to other students" (p.2).

One might expand these individual interpretations to include as well the purposes

(c) to identify strengths and weaknesses and

(d) to measure growth over time.

In addition, there are a variety of

(e) group interpretations of scores that involve score summaries, often averages.

Clearly, the within-grade percentile rank and the within-grade scores derived from it (stanines, NCE's) are designed to give the information in (b) above--where the student stands in relation to other students. Such scores are also useful for (c) to judge strengths and weaknesses but are not designed to give information about (a) what skills a student has or (d) growth over time. Percentile ranks are not useful for group summaries involving averages although other within-grade scores (stanines, NCE's) can be used this way.

Thus, the two types of information the within-grade scores do not provide concern the skills a student has and how those skills are growing over time. It is the desire to provide these two types of information that undoubtedly accounts for the existence of the GE score. This desire to describe what skills a student has has led as well in recent years to criterion referenced testing and reporting. Criterion referenced testing provides information about skills at a very specific level (often the level of a particular instructional objective) whereas GE scores are oriented to a far more global, general level of skills common to a particular grade in school.

GE scores are also directed to information about student growth over time. In the growth area it is useful to differentiate two types of use of growth scores. The simplest is an extrapolation of the type (a) information--to chart what a student knew last year and what he or she knows this year. The second use is to infer not what a student has learned over time but how much. The first type of growth information--the what was learned information--is more related to instruction in schools; the second,--the how much was learned information--to research and evaluation studies. Note that it is the second type of growth interpretation that requires a scale with equal interval properties and it is likely this need that led to the provision by test publishers of various expanded score scales with more nearly equal interval properties. As noted in the criticisms, GE scores are not good for answering the "how much" question about growth but might, if they did provide information about the grade level of performance, help with the "what" question of growth. By contrast most expanded score scales are designed

to help with the "how much" question but the numerical scale carries no "what" meaning.

In my judgment, the first two categories of criticism of GE scores noted above (not equal interval, not comparable across school subjects) are not fundamental criticisms of GE scores at all but statements that the properties of GE scores are not suited to certain uses. (Namely, (b) measuring standing in a group, (c) comparing strengths and weaknesses, the "how much" of (d) measuring growth as well as some types of (e) group summary uses). The concern with equal interval properties stems from evaluation and research efforts to measure growth or individual interpretations of amount gained. As already noted, GE scores are not well suited to these uses and researchers and evaluators should probably be using other scales with better-suited properties (sometimes expanded score scales; sometimes within grade scales) in their studies. Similarly, when comparing a student's or a group's strength in different school subjects, a score reflecting standing within an appropriate grade group (e.g., PR's, stanines, NCE's) should be used, not GE scores.

In my view, the third category of criticism is more fundamental. Even though test makers inform users that the simple desired interpretation is not the one GE scores allow, the fact remains that users wish to know "what skills a student has" (Horst, 1976, p. 2). Further, educators appear to share a common, rough, general idea of the skills and content associated with teaching and learning at each (at least, elementary) grade level to make expression of "what skills" in terms of a grade level designation a meaningful enterprise. The problem is, of course, that GE scores seem now not to be giving this desired and desirable information. There is general agreement that GE scales are

not designed to tell what grade level of work a student can perform (unless the student's performance is near current grade placement) and differences among different GE scales make it clear that they cannot all be providing this information because they are so different. Further, the extreme GE scores presently given very high and very low scorers are just not educationally credible. As noted by Prescott et. al. (1978), to place a third grade pupil receiving a GE of 5.7 in mathematics "in Grade 5 for mathematics instruction could well be a disaster" and "as a rule of thumb: GE's more than 2 years above or below a student's actual grade placement have very little meaning" (p. 23). Since GE scores do not accomplish the purpose of describing what grade level skills a student has, it seems useful to consider if an altered approach might provide this desired information.

A New Grade-Linked Score Scale

Overview of procedure. The goal is to start from the practical constraints of current standardized testing--namely, that the student takes a particular level of a test for which the content range is fairly narrowly confined--and to use performance on this restricted content level to estimate performance on other levels so that a reasonable estimate of the grade level content on which a student can perform can be made.

To accomplish this goal we must have a special study of how students at particular grades perform on content from various grades. Such a study, often called a scaling study, is done in some form to produce present GE score scales, producing data for grade-level combinations such as those shown in Table 1. Presently, however, GE scores are pro-

---

Insert Table 1 about here

---

duced by focusing on the columns and finding the average (or median) performance of each grade on a particular content level and assigning a GE corresponding to that grade to that average score. By contrast, here we propose looking at the rows to see how students at a particular grade perform on various content levels.

The task here is to find for grade X students scoring at a particular score on level X content (the test level corresponding to the student's grade), their scores on other content levels. For example, sixth graders answering 22% of the Level 6 items correctly might be found to answer 46% of Level 5 items, 67% of Level 4, and 76% of Level 3 items. Then in actual use of a Level 6 test with sixth graders, one would estimate that a sixth grader scoring 22% would have, if tested on other levels, scored 46% on Level 5, 67% on Level 4, and 76% on Level 3. This example plus one other is illustrated in Table 2.

Insert Table 2 about here

To convert these estimates of performance on several content levels to a single number describing the level of performance we need adopt a definition of what it means "to perform at grade level X." For the purpose of the Grade Development Scale (GDS) proposed here, performing at grade level X is defined as performing on content designed for grade X (content level X) in the same way students in grade X typically perform on content level X. Thus, from the data described in Table 1, the average (or 50th percentile) performance of grade K students on content level K, grade 1 students on content level 1 through grade 12 students on content level 12 define performance at each of the 13 grade levels. For example, as shown in Table 2 if fourth graders tested at midyear average 67% on the

Level 4 test, then any student estimated to achieve 67% correct on the Level 4 test is described as "performing at the 4.5 grade level" and is given a grade development scale score (GDS score) of 4.5. Similarly, any student estimated to score 65% on Level 5 content is given a GDS score of 5.5. Consequently, the hypothetical sixth grader who scored 22% on Level 6 and was estimated to score 67% on Level 4 would be given a GDS score of 4.5. This GDS score of 4.5 would indicate that the student is estimated to perform on Level 4 content the same as the typical fourth grader performs on Level 4 content--or, in simpler terms, this student is estimated to be performing "at the midyear fourth grade level." Similarly, assuming the hypothetical data in Table 2, a sixth grader scoring 39% correct on the Level 6 test would receive a GDS score of 5.5 indicating that such a student is estimated to perform on Level 5 content as midyear fifth graders do.

Relation of GDS scores and GE scores. The details, complications, and practicalities of implementing this type of grade development scale on a new K-12 achievement test battery, The 3R's Test (Cole, Trent, and Wadell, 1982), are described in Cole (1982). Here we move directly to the results of the procedure and how the GDS scores produced relate to traditional GE scores. First, it must be noted again that although people act as though there is one GE scale on which all tests place scores, it is clear from the different characteristics associated with the GE scales of different tests that different test content, data collection designs, and data analysis procedures have, in fact, produced many different GE scales.

Although various test batteries differ in content and likely in the nature of the norm group on which they are standardized, it is nevertheless

possible to reconstruct the relationship of percentile rank within the national norm group to GE score for any published battery. Such relationships can then be compared to the percentile rank to GDS score relationships found for the GDS score produced here.

Table 3 gives selected percentile rank (PR) to grade score (GDS or GE) conversions for two subjects (Reading and Mathematics) for two grades (4 and 8) for several major achievement batteries. Some of these data are plotted for selected tests in Figure 2 to show the relationship visually. Several observations can be made about the results in this table that are equally true for other grades and other subjects. First, as already noted, GE scales from one test battery are not especially similar to GE scales from other batteries although there is a tendency for scales from the same publisher to be more alike than scales across publishers except for the Metropolitan and Stanford at low grades (probably reflecting different scale construction procedures in use by different publishers).

Second, the GDS scale is consistently different from the GE scales in providing less extreme GDS scores for extreme PR values. That is, an extremely high or low scoring student in a particular grade would be given a GDS score closer to the current grade than any of the GE scores. This is in line with expectations since, as noted earlier, one of the major criticisms of GE scores from practitioners is that they are too extremely different from current grade placement and educationally unrealistic in being so. Apparently, when we estimate directly the level at which students can perform, this unrealistic nature of GE scores is, at least in part, corrected.

Third, the test battery with a GE scale most like the new GDS is

the Iowa Tests of Basic Skills which used a scaling study similar to the one described here but different data analysis procedures to produce its GE's. However, several aspects of the ITBS procedure were implemented in the GDS procedure used here and, in addition, the ITBS procedures were guided by a growth model apparently relatively close to the empirical results from the GDS procedure. Thus, again there seem to be procedural differences as well as different theoretical approaches that produce wide differences in the nature of GE scales produced and that affect how different the scales are from the GDS procedure described here.

Summary

In summary, there seem to be many commonly agreed upon criticisms of grade equivalent (GE) scores. They are not equal interval scales; they should not be used for comparisons across school subjects; and they do not indicate what grade level skills students have. However, when we look at various desired uses of test scores and the types of score scales designed for each use, we find the first two sets of criticisms primarily imply that other scores than GE's should be used for some uses for which GE's are not suited. Some potential uses for GE scores for which other available scores are not suited remain--the use to identify what grade level skills students have and the use to chart growth over time in the "what skills" sense not the "how much growth" sense.

Given these potentially valuable uses, it seemed desirable to consider a variant of a GE scale that could more nearly describe what grade level skills students have. To this end, a new procedure producing a so-called Grade Development Scale (GDS) score was introduced, and some results for the new GDS scores were compared to existing GE scales of

several common tests. The new GDS scale procedure, based on the idea of trying to estimate more directly the grade level of work students can perform, produced less discrepant estimates of grade level from current grade for extreme scores than any of the common GE scales. Thus, it appears possible to avoid one of the pitfalls of GE scores (namely, giving scores that are unrealistically discrepant from current grade) through this new procedure although it remains to be seen if the GDS scores will actually be found useful to educators in practice in estimating the grade level work a student is performing. Studies of this issue are underway.

Whether the GDS procedure as described here proves useful or not, it is clear that the use of GDS scores will not solve some of the problems associated with present GE scores. The scale is not an equal interval scale nor should it be used in comparing subject areas. It may be more appropriate to judge at what grade level students perform, but problems of interpreting the amount of growth in GDS units remains. Also, we still need much more consideration given to meaningful procedures to determine when a student is far enough behind in schoolwork to indicate the need for special attention (likely within-grade national PR's will prove the right solution) and to which type score is most appropriate for examining amount of growth in school for research purposes (an expanded standard score or a within-grade standard score).

Thus, GDS scores are not proposed here as a panacea. Instead the purpose was to sort through the criticisms of GE scores to try to separate those that require switching to a different existing type of score with better properties for a particular use from those that are a more fundamental problem for GE scores. When we did this sorting and isolated

the fundamental thing it would seem usefulfor GE-type scores to do, a way to do this possibly better became apparent. Perhaps this analysis will bring us closer to answering the question of when to GE (for what uses GE type scores are appropriate) and how to GE (how to bring current GE scales closer to the type of information it appears educators desire).

References

Angoff, W. H. Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational measurement. Washington, D.C.: American Council of Education, 1971, pp. 508-600.

Berk, R. A. What's wrong with using grade-equivalent scores to identify LD children? Academic Therapy, 1981, 17, 2, 133-140.

Cole, N. S. A new look at grade-linked score scales. Unpublished manuscript, March, 1982.

Cole, N. S., Trent, E. R., & Wadell, D. C. The 3R's Test. Achievement Edition. (Grades K-12) Chicago, Ill.: The Riverside Publishing Co., 1982.

Hieronymus, A. N., Lindquist, E. F., & Hoover, H. D. Teachers guide, Iowa Test of Basic Skills, (Forms 7/8). Boston: Houghton-Mifflin, 1978.

Horst, D. P. What's bad about grade-equivalent scores. ESEA Title 1 evaluation and reporting system. Technical Report No. 1. Mountain View, CA: RMC Research Corporation, 1976.

Lindvall, C. M., & Nitko, A. J. Measuring pupil achievement and aptitude. New York: Harcourt Brace Jovanovich, Inc., 1975.

Linn, R. L. Measuring pretest-posttest performance changes. In R. A. Berk (Ed.), Educational evaluation methodology: The state of the art. Baltimore: The Johns Hopkins University Press, 1981, pp. 84-109.

Linn, R. L., & Slinde, J. A. The determination of the significance of change between pre- and posttesting periods. Review of Educational Research, 1977, 47, 1, 121-150.

Prescott, G. A., Balow, I. H., Hogan, T. P., & Carr, R. C. Metropolitan Achievement Tests Teachers manual for administering and interpreting. Forms JS and KS. New York: Harcourt Brace Jovanovich, 1978.

Reynolds, C. R. The fallacy of "two years below grade level for age" as a diagnostic criterion for reading disorders. Journal of School Psychology, 1981, 19, 4, 350-358.

Thorndike, R. L., & Hagen, E. P. Measurement and evaluation in psychology and education. New York: John Wiley & Sons, 1977.

Table 1

Hypothetical Scaling Study Design

Student	Content Level											
Grade	K	1	2	3	4	5	6	7	8	9	10	11	12
K	x	x	x	x									
1	x	x	x	x	x								
2	x	x	x	x	x	x							
3	x	x	x	x	x	x	x						
4		x	x	x	x	x	x	x					
5			x	x	x	x	x	x	x				
6				x	x	x	x	x	x	x			
7					x	x	x	x	x	x	x	x	
8						x	x	x	x	x	x	x	x
9							x	x	x	x	x	x	x
10								x	x	x	x	x	x
11									x	x	x	x	x
12										x	x	x	x

Table 2

Some Hypothetical Data

Percent Correct for Sixth Grade Students

	Scores on Other Levels for Students Scoring at the Designated Percent Correct on Level 6 Content		
Level 6	Level 5	Level 4	Level 3
22%	46%	67%	76%
39%	65%	73%	81%

Percent Correct Performance "At Grade Level"

Grade of Students	Content Level	Percent Correct
3.5	3	64%
4.5	4	67%
5.5	5	65%

Table 3

Comparisons of GDS Scales and GE Scales for Selected Achievement Batteries

Percentile Rank	The 3R's GDS	CAT[a] GE	CTBS[b] GE	ITBS[c] GE	Metro[d] GE	SAT[e] GE
		Grade 4 - Reading - Spring				
98	7.1	10.0	10.0	80	12.1	9.5+
95	6.9	8.6	9.1	75	10.8	9.3
90	6.6	7.8	8.1	70	9.7	8.1
75	5.7	6.2	6.3	60	7.3	6.6
50	4.7	4.7	4.7	48	4.7	5.0
25	3.9	3.6	3.5	37	3.1	3.6
10	3.2	2.7	2.5	29	2.4	2.7
5	3.0	2.4	2.1	25	2.2	2.2
2	2.8	1.8	1.8	21	2.0	2.0
		Grade 4 - Mathematics - Spring				
98	5.9	7.8	8.7	70	10.9	8.6
95	5.8	7.2	7.7	67	8.7	7.7
90	5.6	6.6	6.9	63	7.5	6.8
75	5.3	5.6	5.7	56	6.2	6.0
50	4.7	4.7	4.7	48	4.9	5.0
25	4.2	3.8	3.5	40	3.7	3.9
10	3.9	3.3	2.8	35	2.8	3.0
5	3.7	2.8	2.4	32	2.4	2.6
2	3.6	2.3	2.0	29	2.1	2.0
		Grade 8 - Reading - Spring				
98	12.0	12.9+	13.6	127	12.9+	12.8
95	11.7	12.9+	13.6	119	12.9+	12.8
90	11.3	12.9+	13.2	112	12.9+	12.8
75	10.2	11.1	10.8	102	12.4	11.5
50	8.7	8.7	8.7	88	9.0	9.1
25	7.4	6.7	6.2	70	6.5	6.7
10	6.3	4.9	4.4	57	4.2	4.7
5	6.0	4.1	3.5	50	3.4	3.6
2	5.7	3.4	2.8	43	2.8	3.0
		Grade 8 - Math - Spring				
98	11.2	12.5+	13.6	117	12.9+	12.6
95	11.0	12.5+	13.6	112	12.9+	12.6
90	10.7	12.5+	13.6	107	12.9+	12.6
75	9.9	12.5+	10.5	98	12.9+	11.7
50	8.7	8.7	8.7	87	9.0	9.0
25	7.5	7.0	6.3	74	6.9	6.8
10	6.8	5.5	4.7	64	5.7	5.5
5	6.5	4.8	3.7	59	4.9	4.7
2	6.3	3.8	2.9	54	4.1	4.0

[a] California Achievement Tests Norms Tables, Levels 14C and 18C. Monterey, CA: CTB/McGraw-Hill, 1978.

[b] Comprehensive Tests of Basic Skills Examiner's Manual. Levels 2 and 4. Monterey, CA: CTB/McGraw-Hill, 1974.

[c] Iowa Tests of Basic Skills Teacher's Guide. Iowa City, IA: Houghton Mifflin, 1979.

[d] Metropolitan Achievement Tests Teacher's Manual for Administering and Reporting, Elementary and Advanced 1. New York: The Psychological Corporation, 1978.

[e] Stanford Achievement Test Norms Booklet, Intermediate 1 and Advanced. New York: Harcourt Brace Jovanovich, 1973.

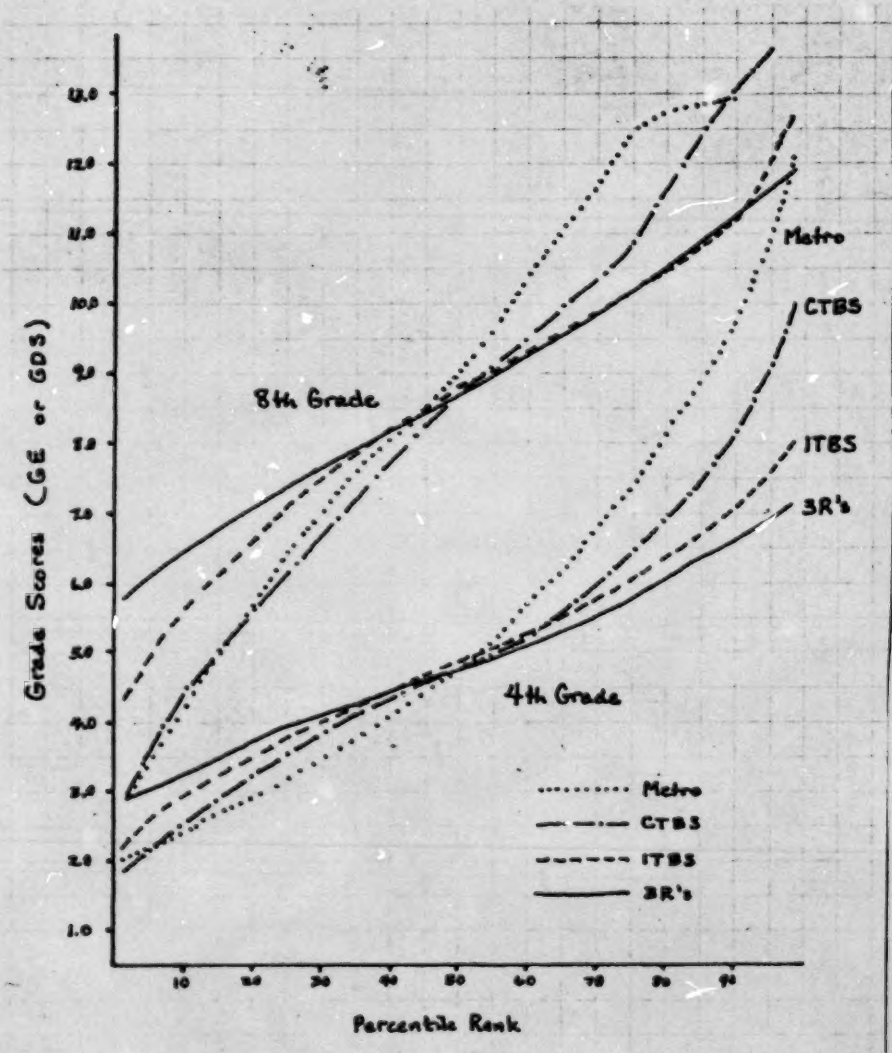

Figure 1. Percentile rank vs. grade score plots for selected tests for grades 4 and 8, Spring, in Reading for data taken from Table 2.